R vs. Python: for Data Analysis

Katherine Encarnacion

MATG 611

December 18, 2015

Introcuction

▶ Using Python to Reproduce midterm 2

▶ What is Python?

▶ What is R?

Data Manipulation

Python Packages and Libraries

- ► Importance of libraries and packages in Python
- ► NumPy
- Pandas
- CSV
- matplotlib

```
import pandas as pd
import numpy as np
import csv as csv
import matplotlib.pyplot as plt
```

Figure: Imported libraries and Packages

Problem 1

Find the mean of the heart rate of each subject

Simple syntax

Errors within the book

► Easier to learn then R

Problem 1

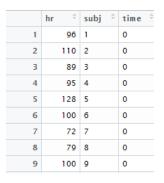


Figure: Problem 1

Incorrect Syntax

Figure: Python for Data Analysis Book

R Results for Problem 1

	heart.rate.tapply $^{\diamondsuit}$
1	91.50
2	109.50
3	85.75
4	83.50
5	122.00
6	98.00
7	69.50
8	75.50
9	103.00

Figure: Problem 1 Part a in R

Python Results for Problem 1

	hr
subj	
1	91.50
2	109.50
3	85.75
4	83.50
5	122.00
6	98.00
7	69.50
8	75.50
9	103.00

Figure: Problem 1 Part a in Python

Problem 2

Creating factors

Creating levels

Similar Syntax to R

Used counting function instead of creating a table

Python Syntax for Problem 2

group = pd.cut(blood,bins,labels=group_name)

Figure: Problem 2 in Python using Cut function

R Syntax for Problem 2

```
grouped Blood <- \ cut(blood,x,labels = c("low","intermediate","high","very \ high"))
```

Figure: Problem 2 in R using the Cut function

Conclusion

▶ It is quite difficult to learn new syntax in 2 days.

Both are very similar.

Difficulty solving problem 3.

Prefer R for data analysis.